

Warm Starting Series of Mixed-Integer Linear Programs

Sean Kelley ^a

^a Department of Industrial and Systems Engineering, Lehigh University

January 3, 2022

Abstract

When one solves a mixed-integer linear programming (MILP) problem with today’s state-of-the-art solvers, the solvers employ a Branch and Cut algorithm that returns an optimal solution while collecting a variety of additional information embedded within a binary tree. The information in this tree includes other feasible solutions, valid inequalities derived from cut generation, and a disjunction of the variables, and it is collected to close the primal-dual gap, the terminating condition of the algorithm. When one solves another MILP problem that differs from the one before only by objective coefficients or bounds on the constraints, some attributes of the previous problem’s tree remain valid in the new one or are easily translated. By reusing this information, either directly or by using it to generate valid inequalities on the initial LP relaxation, we can tighten the primal and dual bounds more effectively than by the solver’s primal heuristics alone, leading to a quicker algorithm execution. Such conditions to reuse a previous tree may seem restrictive, but several well-known classes of mixed-integer programming (MIP) problems rely on finding solutions by solving a series of sufficiently structured problems. In this work, we explore different means of reusing information from previous MILP solves and explore the effectiveness of doing so in improving the efficiency of solving more complex classes of MIP problems.

1 Introduction

1.1 Motivation

Mixed-Integer Programming (MIP) is a class of problems in which one minimizes an objective subject to a set of constraints and the restriction that some variables must take on integer values. MIP has a deeply studied theory and has been adopted in many successful industrial applications, such as production scheduling, vehicle routing, and facility location. Much of this theory and many of these applications are focused on the case in which a single Mixed-Integer Linear Programming (MILP) instance, which is a MIP instance with linear constraints and objective, is solved. In recent decades, this theory has been built upon to enrich the space of solvable MIP instances. Some examples include solving decomposable MILP’s with many constraints or variables, MILP’s with multiple objectives, MIP’s with nonlinear objective or constraints (MINLP), and MILP’s with ”real-time” run time restrictions. With this growth in theory, MIP is more relevant to industry than ever, specifically in the aforementioned applications.

A major bottleneck to implementing applications of more recently developed subclasses of MIP, like those mentioned above, is solving them in a reasonable amount of time. The computational difficulty in the above examples stems from the fact that each relies on solving a series of many MILP instances, where solving a single instance is already an NP-hard problem. However, in these cases, each MILP instance in the series shares a similar structure with the others in that they share coefficient matrices and differ only in right hand sides (RHS’s) or objectives. In this paper, we show how this structure can be taken advantage of to greatly improve the run time when solving many such instances in a row.

1.2 Approach

In this work, we improve the run time of each MILP solve in such closely related series (referred to here on out as a "family") by "warm starting" its solver, which means that we provide the solver with information beyond the model formulation that shrinks the space in which it will have to search for a solution. Specifically, this means we provide the solver with information that reduces the primal-dual gap, which we do in the following four ways:

1. We provide the solver with feasible solutions found by previous MILP solves in the family. When only the objective differs among the instances, we can set the primal bound of the current instance to the objective value of the best previously found solution in terms of the current objective.
2. We provide the solver with a dual function generated by previous MILP solves in the family. When only the RHS differs among the instances, we can set the dual bound of each LP relaxation in the current instance's branch and bound tree to the evaluation of the dual function at the relaxation's right hand side.
3. We provide the solver with valid inequalities generated by previous MILP solves in the family. When only the objective differs amongst the instances, we can add all of the previously valid inequalities to the current instance as they all share the same feasible region. When the RHS's change amongst the instances, we can re-evaluate the methods for generating each previously valid inequality at the current RHS to yield inequalities valid for the current instance.
4. We provide the solver with a disjunction generated by previous MILP solves in the family. We can input this disjunction into a cut generating linear program (CGLP) that will generate valid cuts during each cut-generating routine of branch and cut, which refines the feasible region near the optimal solution.

By reducing the primal-dual gap via the above four methods, we reduce the number of nodes in the branch and bound tree that we must bound before finding the optimal solution. We demonstrate in this paper the effectiveness of doing so against a sample of families in each of the seven following problem types: restarting single MILP instances, dual decomposition for stochastic MILP, Branch and Price, Multi-Objective MILP, MINLP, real-time MILP, and serialized Local Search primal heuristic.

1.3 Outline

This paper is organized as follows. In section 2 we cover preliminaries on MILP, and in section 3 we cover the details on what each numerical experiment will have in common. We dedicate sections 4 to 10 to each of the six problem types above. Each of those sections covers how the given problem type builds upon MILP, including what is the MILP family from which many similar instances will be solved, how we incorporate the four aforementioned warm starting methods, and how our warm started approach compares to solving each MILP. We conclude in section 11 with our thoughts on where warm starting MILP instances is most appropriate and further research directions.

1.4 Literature Review

This work is largely an extension of [5], which warm-starts MILP's differing only by objective or constraint bounds. In that work, warm-starting is achieved by reusing the Branch-and-Cut tree and cuts generated during the solve process. The main drawback to that approach, however, is that the tree can be large, requiring the solver to explore many possible paths to the optimal solution. Warm starting by only adding strong cuts for the new MILP instance to the root relaxation circumvents this issue, preserving a refined relaxation while leaving only one node to search instead of many. The common approach to find these cuts is through Lift-and-Project [2]. Until recently, such an approach was considered intractible when using a Branch-and-Cut tree with more than a few disjunctive terms, but [1] proposes a new formulation for the problem, enabling its application at scale. Our contribution is to generate cuts in this fashion from a

previously solved MILP Branch-and-Cut tree, using them alone to warm start a new MILP instance and gaining the aforementioned advantages.

Before continuing, it is worth comparing our approach to the learning-based approaches to warm starting, which have soared in popularity over the last few years. Learning-based approaches train neural networks to predict primal heuristic and branching decisions made during branch and bound algorithms [4], [3]. When put into practice, such approaches enable solvers to more quickly navigate the branch and bound tree for new instances. This differs from our approach as we improve the solve time of a MILP instance by reducing the space in which a solver must search for a solution. One advantage of learning-based warm-starts is that they give the flexibility that not every problem learned on must have as similar structure as the problems we are considering for predictions to be accurate. For example, the problems could semantically be the same but differ in number of constraints and variables. A disadvantage, however, is that learning-based approaches require many computational and human resources to generate warm-starts, which must be repeated when applied to a new problem. On the other hand, the framework we propose for sufficiently structured problems requires relatively little computational overhead and can be quickly retrofitted to new applications. It is worth noting, though, that because both approaches work in different ways to improve the run time of MILP solves, it could be worth exploring the effectiveness of using both in tandem in the future.

2 Preliminary Material

In this section we present background concepts that we reference and build upon in the remainder of the paper.

2.1 Mixed Integer Linear Programming

The most fundamental concept to this paper is that of the Mixed Integer Linear Program, which is defined as follows:

$$\begin{aligned} & \text{minimize } c^T x \\ & \text{subject to } Ax \geq b \\ & \quad x \geq 0 \\ & \quad x_i \in \mathbb{Z} \text{ for all } i \in \mathcal{I} \end{aligned} \tag{1}$$

In the above, x is our variable and \mathcal{I} is the set of indices of x that must take on integer values. We assume $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$.

(1) has a linear programming (LP) relaxation, which is defined the same as above, except the integer restriction on the indices belonging to \mathcal{I} is removed. The LP relaxation is defined as follows:

$$\begin{aligned} & \text{minimize } c^T x \\ & \text{subject to } Ax \geq b \\ & \quad x \geq 0 \end{aligned} \tag{2}$$

The solution to (2) is a lower bound on (1) since removing constraints expands the solution set.

A common approach to solving MILP's is known as *Branch and Bound*. There are many resources that give the full details on the execution of this class of algorithm. The details relevant to our work are as follows. First, "bound" (1) by solving (2), since the latter can be done efficiently. Let x^* be the solution to (2) and $i \in \mathcal{I}$ be an index of x^* violating the integrality requirement. Next, "branch" by creating two new instances of (2) where one has the additional restriction that $x_i \leq \lfloor x_i^* \rfloor$ and the other $x_i \geq \lceil x_i^* \rceil$. Then recursively repeat this process on each new instance until a solution satisfying the integrality constraint in (1) is found or we "prune" the instance by deciding to not solve it. One of the most often used subclasses of Branch and Bound used in practice is known as *Branch and Cut*. It follows a similar procedure to Branch and Bound,

except between bounding and branching an additional routine is run to generate additional *cuts*, which are inequalities valid for (1) that refine the current instance's feasible region.

Each of these instances can be represented as a node, t , in a binary tree where edges represent the additional restriction on x added from solving the parent node's instance. We define each node's ancestors as the list of nodes connecting it to the root (including itself and the root), where the root node represents the instance (2). We let $a(t)$ be the map of a node to its ancestors. For a given node t in our Branch and Bound binary tree, we define u^t as the upper bound restrictions and l^t as the lower bound restrictions placed on x as a result of branching. If no upper bound or lower bound restriction has been placed on index $i \in [n]$ of x in node t , then $u_i^t = \infty$ and $l_i^t = 0$, respectively. Then the LP instance corresponding to node t can be represented as follows:

$$\begin{aligned} & \text{minimize } c^T x \\ & \text{subject to } Ax \geq b \\ & \quad u^t \geq x \geq l^t \end{aligned} \tag{3}$$

When a solution to (3) for some node t is integer, it is a solution to (1). As we execute Branch and Bound, we refer to the best known solution to (1) as the *primal bound* on (1).

Consider a set of nodes such that the union of the feasible regions of their LP instances contains the feasible region to 1 but their intersection is empty. We refer to the collection of branching decisions that yielded this set of nodes as a *disjunction*.

One important disjunction to note is that which corresponds to the terminal nodes of the Branch and Bound tree. Let this set of nodes be represented by \mathcal{T} . Since \mathcal{T} corresponds to a disjunction, all possible solutions to (1) belong to the union of the feasible regions of each disjunctive term's LP. Therefore, the minimum across each terminal node's LP's optimal objective value is a lower bound on (1). We call this lower bound the *dual bound* on (1).

When monitoring the progress of a MILP solver, we can compare how close the best optimal objective value we can possibly find for a solution to the best objective value we have already found for a solution. This is often referred to as the *primal-dual gap* and is defined as follows:

$$\text{primal-dual gap} = \frac{\text{primal bound} - \text{dual bound}}{\text{primal bound}} \tag{4}$$

2.2 Cut Generating Linear Program

One type of cut we use frequently in this work during the cut generation routine of Branch and Cut are cuts generated from the Cut Generating Linear Program (CGLP). We can derive these cuts as follows.

Let t be a node in a branch and bound tree, $\mu^t \in \mathbb{R}_+^m$, $w^t \in \mathbb{R}_+^n$, $v^t \in \mathbb{R}_+^n$, and I_n be the n -dimensional identity matrix. Then we know

$$A^T \mu^t + I_n w^t - I_n v^t \geq b^T \mu^t + l^t w^t - u^t v^t \tag{5}$$

is a valid inequality for (3).

Recall the set of terminal nodes, \mathcal{T} , of a Branch and Bound tree. Let π and π_0 be such that the following holds:

$$\begin{aligned} \pi & \geq A^T \mu^t + I_n w^t - I_n v^t \\ \pi_0 & \leq b^T \mu^t + l^t w^t - u^t v^t \end{aligned} \quad \text{for all } t \in \mathcal{T} \tag{6}$$

Since (5) is a valid inequality for each LP in the disjunction induced by \mathcal{T} , it follows $\pi^T x \geq \pi_0$ is a valid inequality for each LP in the disjunction induced by \mathcal{T} . Inequalities that are valid for each underlying LP in a disjunction are referred to as *disjunctive inequalities*.

Often times the goal with disjunctive inequalities is to cut off an optimal solution to (3) that violates the integrality constraint in (1). In doing so, we aim to tighten the dual bound by as much as we can, which

means to find a valid cut for (1) such that the solution to (3) is violated by as much as possible. Given a solution to (3), x^* , this yields the following LP formulation:

$$\begin{aligned}
& \text{minimize} && \pi^T x^* - \pi_0 \\
& \text{subject to} && \pi \geq A^T \mu^t + I_n w^t - I_n v^t && \text{for all } t \in \mathcal{T} \\
& && \pi_0 \leq b^T \mu^t + l^t w^t - u^t v^t && \text{for all } t \in \mathcal{T} \\
& && 1 = \sum_{t \in \mathcal{T}} \left(\sum_{j=1}^m \mu_j^t + \sum_{i=1}^n w_i^t + \sum_{i=1}^n v_i^t \right) \\
& && \mu^t \in \mathbb{R}_+^m, w^t \in \mathbb{R}_+^n, v^t \in \mathbb{R}_+^n && \text{for all } t \in \mathcal{T}
\end{aligned} \tag{7}$$

This formulation is valid for the following reasons. (6) implies that (7) generates a valid disjunctive inequality. Since $\pi^T x \geq \pi_0$ is valid for (1), the objective of (7) finds the most violated disjunctive inequality since (7) has a normalization term in its constraints.

Although (7) finds a deep disjunctive inequality, it requires many multiples of the coefficient matrix of (1), which makes it intractible for disjunctions of more than a few terms. [1] proposes a method of generating deep disjunctive inequalities where for each terminal node, t , of a subtree, A and b in (7) are replaced by the subset of constraints that are tight at the solution to t 's LP instance. In a future edition of this work, we will derive this revised CGLP and implement it for numerical experiments. By reducing the numbers of variables and constraints in the problem, however, we are able to make the CGLP a much more tractible model in practice for cut generation.

Note, now that we have found valid cuts for (1) that are added to the LP in node t , we need to update its formulation. Let Π^t and Π_0^t represent the coefficient matrix and RHS of cuts added to the LP instance in node t . Then its formulation is as follows:

$$\begin{aligned}
& \text{minimize} && c^T x \\
& \text{subject to} && Ax \geq b \\
& && \Pi^t x \geq \Pi_0^t \\
& && u^t \geq x \geq l^t
\end{aligned} \tag{8}$$

2.3 Re-Evaluating Cuts When Right Hand Side Changes

One significant challenge to warm starting when one recycles cuts from previously solved instances to a new one is ensuring the cuts stay valid if the RHS of (1) changes. Fortunately for this publication, the two cuts we employed, Gomory Mixed Integer Cuts (GMICs) and CGLP cuts, both are derived from the value of the RHS. Given we know their derivation, we can deduce quickly for a new RHS how each type of cut should be updated to remain valid.

We begin by looking at how to update cuts from the CGLP. To ensure (π, π_0) , a CGLP cut generated after bounding node t , remain valid for a new RHS, β , we look at (6) to see what conditions must hold. We define $p_t(\beta)$ as the map such that $(\pi, p_t(\beta))$ remains valid. Thus, we define $p_t(\beta)$ as follows:

$$p_t(\beta) = \min\{\beta^T \mu^t + l^t w^t - u^t v^t \text{ for all } t \in \mathcal{T}\} \tag{9}$$

Note in the above that for all $t \in \mathcal{T}$, μ^t , w^t , v^t retain their values from the solution of (7) after bounding node t . If we wanted to be a little more exact, we could put β in place of b in (7) and solve it again, starting from its previous optimal basis. This would deliver a deeper cut at the expense that it would take longer to generate.

We also can follow a similar idea to update GMICs when the RHS changes. In a future version of this work, we will detail that procedure as well.

Moving forward, let $P_t(\beta)$ be a map that updates the set of cuts added to the LP in node t such that all cuts remain valid when we change the RHS of (1) to β . Each dimension of the map represents the appropriate transformation above.

2.4 Mixed Integer Linear Programming Duality

As mentioned in Section 1.2, we create a dual of (1) as a function of its RHS, β . We derive this dual function as follows.

For a new RHS, β , in the LP for node t , we can update the cut matrix as mentioned section 2.3 with the map $P_t(\beta)$. Then building off of (8), we get an updated formulation for the LP in node t , which is as follows:

$$\begin{aligned} & \text{minimize } c^T x \\ & \text{subject to } Ax \geq \beta \\ & \quad \Pi^t x \geq P_t(\beta) \\ & \quad u^t \geq x \geq l^t \end{aligned} \tag{10}$$

By LP duality theory, we know we can obtain a lower bound on its objective value of (10) via its dual, which has the following formulation:

$$\begin{aligned} & \text{maximize } \beta^T \delta^t + P_t(\beta)^T \delta_{\Pi}^t + l^{t^T} \underline{\delta}^t + u^{t^T} \bar{\delta}^t \\ & \text{subject to } A^T \delta^t + \Pi^{t^T} \delta_{\Pi}^t + \underline{\delta}^t + \bar{\delta}^t = c \\ & \quad \delta^t, \delta_{\Pi}^t, \underline{\delta}^t \geq 0; \quad \bar{\delta}^t \leq 0 \end{aligned} \tag{11}$$

By the definition of the dual bound in section 2.1, we know the minimum of (10) across all $t \in \mathcal{T}$ is a lower bound on (1) evaluated at a new RHS β . Therefore, the minimum of (11) is such a lower bound as well. Further, since the objective of (11) evaluated at any of its feasible solutions is a lower bound on (11), the minimum of the objective of (11) evaluated at the optimal solution $(\delta^t, \delta_{\Pi}^t, \underline{\delta}^t, \bar{\delta}^t)$ for RHS b for each $t \in T$ is a lower bound on (1) evaluated at a new RHS β . Therefore, we have the following dual function:

$$\underline{\phi}^T(\beta) = \min_{t \in \mathcal{T}} \{ \beta^T \delta^t + P_t(\beta)^T \delta_{\Pi}^t + l^{t^T} \underline{\delta}^t + u^{t^T} \bar{\delta}^t \} \tag{12}$$

Since LP's remain dual feasible when constraints are added to the primal, the previous solutions to all of node t 's ancestry, defined as $a(t)$ in section 2.1, are all lower bounds to (11). Thus we can tighten (12) as follows:

$$\underline{\phi}^T(\beta) = \min_{t \in \mathcal{T}} \{ \max_{t' \in a(t)} \{ \beta^T \delta^{t'} + P_{t'}(\beta)^T \delta_{\Pi}^{t'} + l^{t'^T} \underline{\delta}^{t'} + u^{t'^T} \bar{\delta}^{t'} \} \} \tag{13}$$

3 General Evaluation

Although we are running numerical experiments against a variety of problem classes, there are a few parts of our experiments that will remain constant through out our paper. First, we run each MILP instance until the primal-dual gap is 1%. We take this approach since we want to understand the effects our work will have on problems in practice, and this is a common tolerance to use in practice. We then have two metrics we use to compare the effectiveness of our warm starting approach to cold starting. One attribute is the amount of time it takes to achieve 1% primal-dual gap, as that tells us clearly whether warm starting is currently more effective than cold starting. The other attribute we track is number of node evaluations to 1% primal-dual gap. This tells us if it is possible to achieve a better total run time with warm starting versus cold starting if we can collect the warm started information and generate cuts efficiently enough. For each problem class, we will look at the distribution of how warm starting compares to cold starting in both regards over a sample of solves to decide if warm starting is or can be more effective than cold starting.

4 Restarting MILP's

The first problem class we examine is restarting a MILP instance after a partial solve so that we can warm start it with information discovered about the instance in the initial solve. The hope here is that the

performance improvement we receive in the warm started solve will offset the effort taken in the initial solve so that both together are less work than a typical cold start. We look to identify which MILP instances, if any, in which this happens. Since we only have a single, partial solve with which we warm start, and the partial solve counts in the time and evaluations taken, we do not expect much here. However, we test this one as a success here would have far reaching implications on how most Branch and Cut solvers work today.

4.1 Extension from Single Solve MILP

As mentioned above, this problem class still belongs to MILP, but it differs from the cold started approach to solving MILP's in that we do a partial cold start and then warm start the exact same instance with information generated from the cold start.

4.2 Warm Starting Approach

Since we solve the exact same problem for the warm started instance as we do for the previous cold started instance, we are not restricted in what warm starting information from section 1.2 that we can use. Thus, we use them all.

4.3 Numerical Results

In addition to capturing run time and node evaluations to 1% primal-dual gap, we track the dual gap in the warm started instance after processing the root node and after we process as many nodes as the partial solve (known as our *cut off*). In these cases, we use the dual gap instead of the primal-dual gap since we know what the optimal objective value of the problem is but do not always have a primal bound to compare to when restarting after only a few nodes have been processed. For reference the dual gap is defined as follows:

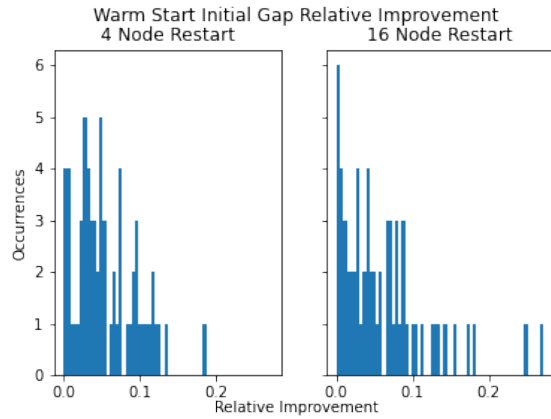
$$\text{dual gap} = \frac{\text{optimal objective value} - \text{dual bound}}{\text{optimal objective value}} \quad (14)$$

We run our experiment as follows. We create 64 random MILP instances, varying in terms of the numbers of constraints, variables, and nonzero constraint coefficients, the range of constraint and objective coefficient values, and how tight each constraint is. Additionally, we vary after how many processed nodes do we restart the instance when warm starting. We then collect the four aforementioned metrics for each run to create a distribution of each.

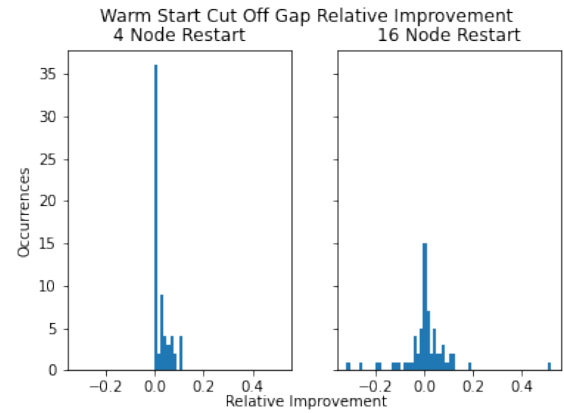
At time of writing, we are still implementing step 4 from our approach in section 1.2 and can further improve each of the first three, so the computational results below are more to give an example of how results will be organized than what they will end up being. Where we stand currently, however, is with the following results:

As we can see in Figure 2, we worsen in both nodes evaluated and run time in most cases. We interpret this as the CGLP cuts we carry over from the initial solve to the warm start is not being helpful for improving the performance. We believe this to be the case as generating many cuts from the CGLP is more likely to be helpful as the extra cuts will further reduce primal-dual gap. We do see some promise in using CGLP cuts in Figure 1 since the single cut we can generate from the initial solve does result in initial improvements in the dual gap. This leads us to believe that having a cut generating method where we could create CGLP cuts after bounding each node might be of great benefit to us. In a future version of this work, we will detail more what improvements we get from doing so.

5 Dual Decomposition for Stochastic MILP's



(a) Relative improvement of warm start dual gap over cold start dual gap after evaluating the root



(b) Relative improvement of warm start dual gap over cold start dual gap after evaluating as many nodes as the restart requirement

Figure 1: Comparisons of the additional metrics

5.1 Extension from Single Solve MILP

5.2 Warm Starting Approach

5.3 Numerical Results

6 Branch and Price

6.1 Extension from Single Solve MILP

6.2 Warm Starting Approach

6.3 Numerical Results

7 Multi-Objective MILP

7.1 Extension from Single Solve MILP

7.2 Warm Starting Approach

7.3 Numerical Results

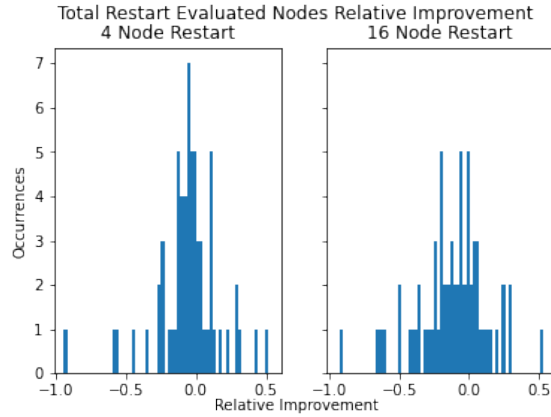
8 MINLP

8.1 Extension from Single Solve MILP

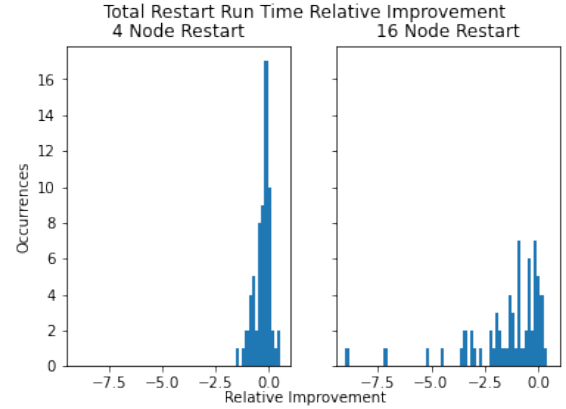
8.2 Warm Starting Approach

8.3 Numerical Results

9 Real-Time MILP



(a) Relative improvement of warm start node evaluations over cold start node evaluations to reach 1% primal-dual gap



(b) Relative improvement of warm start run time over cold start run time to reach 1% primal-dual gap

Figure 2: Comparisons of main metrics

9.1 Extension from Single Solve MILP

9.2 Warm Starting Approach

9.3 Numerical Results

10 Serialized Local Search Primal Heuristic

10.1 Extension from Single Solve MILP

10.2 Warm Starting Approach

10.3 Numerical Results

11 Conclusion

As it stands currently, we need to finish developing working versions of all four warm starting techniques we listed in section 1.2, as having cut generating methods after bounding a node is still a work in progress. Further, we need to optimize each technique to be efficient to give us the best comparison to cold start run times as we can get. Lastly, we then need to implement a solve method for what remaining problem subclasses we want to test and repeat the experiment outlined in section 3. After doing that, we hope to conclude with how effective warm starting is for each problem subclass we test and what further research directions remain.

References

- [1] Balas and Kazachov. V-polyhedral disjunctive cuts. *N/A*, 2021.
- [2] Sebastián Ceria Egon Balas and Gérard Cornuéjols. A lift-and-project cutting plane algorithm for mixed 0-1 programs. *Math. Program.*, 1993.
- [3] Lodi et. al. Learning to schedule heuristics in branch-and-bound. *ArXiv*, 2019. doi:<https://arxiv.org/abs/2103.10294>.
- [4] Nair et. al. Solving mixed integer programs using neural networks. *ArXiv*, 2020. doi:<https://arxiv.org/pdf/math/0411298v1.pdf>.
- [5] Ralphs. Warm starting mixed integer programs. *COR@L*, 2006. doi:<https://coral.ise.lehigh.edu/~ted/files/papers/DMII06.pdf>.